

10950

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: CAPUT, DANIEL
FERRARA, PASCUAL
GUILLEMOT, JEAN-CLAUDE
KAGHAD, MOURAD
LEGOUX, RICHARD
LOISON, GERARD
LARBRE, ELIZABETH
LUPKER, JOHANNES
LEPLATOIS, PASCUAL
SALOME, MARK



(ii) TITLE OF INVENTION: URATE OXIDASE ACTIVITY PROTEIN,
RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
MICRO-ORGANISMS AND TRANSFORMED CELLS

(iii) NUMBER OF SEQUENCES: 36

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Foley & Lardner
(B) STREET: 1800 Diagonal Road, Suite 500
(C) CITY: Alexandria
(D) STATE: Virginia
(E) COUNTRY: USA
(F) ZIP: 22313-0299

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/659,408
(B) FILING DATE: 25-APR-1991
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: BENT, Stephen A.
(B) REGISTRATION NUMBER: 29,768
(C) REFERENCE/DOCKET NUMBER: 16781/276 BEDL

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (703)836-9300
(B) TELEFAX: (703)683-4109
(C) TELEX: 899149(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 301 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Aspergillus flavus

(vii) IMMEDIATE SOURCE:

(B) CLONE: Urate oxidase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Ser Ala Val Lys Ala Ala Arg Tyr Gly Lys Asp Asn Val Arg Val Tyr
1 5 10 15
Lys Val His Lys Asp Glu Lys Thr Gly Val Gln Thr Val Tyr Glu Met
20 25 30
Thr Val Cys Val Leu Leu Glu Gly Glu Ile Glu Thr Ser Tyr Thr Lys
35 40 45
Ala Asp Asn Ser Val Ile Val Ala Thr Asp Ser Ile Lys Asn Thr Ile
50 55 60
Tyr Ile Thr Ala Lys Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly
65 70 75 80
Ser Ile Leu Gly Thr His Phe Ile Glu Lys Tyr Asn His Ile His Ala
85 90 95
Ala His Val Asn Ile Val Cys His Arg Trp Thr Arg Met Asp Ile Asp
100 105 110
Gly Lys Pro His Pro His Ser Phe Ile Arg Asp Ser Glu Glu Lys Arg
115 120 125
Asn Val Gln Val Asp Val Val Glu Gly Lys Gly Ile Asp Ile Lys Ser
130 135 140
Ser Leu Ser Gly Leu Thr Val Leu Lys Ser Thr Asn Ser Gln Phe Trp
145 150 155 160
Gly Phe Leu Arg Asp Glu Tyr Thr Thr Leu Lys Glu Thr Trp Asp Arg
165 170 175
Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys Asn Phe Ser
180 185 190
Gly Leu Gln Glu Val Arg Ser His Val Pro Lys Phe Asp Ala Thr Trp
195 200 205
Ala Thr Ala Arg Glu Val Thr Leu Lys Thr Phe Ala Glu Asp Asn Ser
210 215 220
Ala Ser Val Gln Ala Thr Met Tyr Lys Met Ala Glu Gln Ile Leu Ala
225 230 235 240
Arg Gln Gln Leu Ile Glu Thr Val Glu Tyr Ser Leu Pro Asn Lys His
245 250 255
Tyr Phe Glu Ile Asp Leu Ser Trp His Lys Gly Leu Gln Asn Thr Gly
260 265 270
Lys Asn Ala Glu Val Phe Ala Pro Gln Ser Asp Pro Asn Gly Leu Ile
275 280 285
Lys Cys Thr Val Gly Arg Ser Ser Leu Lys Ser Lys Leu
290 295 300

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 302 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
 (iii) HYPOTHETICAL: NO
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Aspergillus flavus
 (vii) IMMEDIATE SOURCE:
 (B) CLONE: Met-Urate oxidase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ser Ala Val Lys Ala Ala Arg Tyr Gly Lys Asp Asn Val Arg Val
 1 5 10 15
 Tyr Lys Val His Lys Asp Glu Lys Thr Gly Val Gln Thr Val Tyr Glu
 20 25 30
 Met Thr Val Cys Val Leu Leu Glu Gly Glu Ile Glu Thr Ser Tyr Thr
 35 40 45
 Lys Ala Asp Asn Ser Val Ile Val Ala Thr Asp Ser Ile Lys Asn Thr
 50 55 60
 Ile Tyr Ile Thr Ala Lys Gln Asn Pro Val Thr Pro Pro Glu Leu Phe
 65 70 75 80
 Gly Ser Ile Leu Gly Thr His Phe Ile Glu Lys Tyr Asn His Ile His
 85 90 95
 Ala Ala His Val Asn Ile Val Cys His Arg Trp Thr Arg Met Asp Ile
 100 105 110
 Asp Gly Lys Pro His Pro His Ser Phe Ile Arg Asp Ser Glu Glu Lys
 115 120 125
 Arg Asn Val Gln Val Asp Val Val Glu Gly Lys Gly Ile Asp Ile Lys
 130 135 140
 Ser Ser Leu Ser Gly Leu Thr Val Leu Lys Ser Thr Asn Ser Gln Phe
 145 150 155 160
 Trp Gly Phe Leu Arg Asp Glu Tyr Thr Thr Leu Lys Glu Thr Trp Asp
 165 170 175
 Arg Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys Asn Phe
 180 185 190
 Ser Gly Leu Gln Glu Val Arg Ser His Val Pro Lys Phe Asp Ala Thr
 195 200 205
 Trp Ala Thr Ala Arg Glu Val Thr Leu Lys Thr Phe Ala Glu Asp Asn
 210 215 220
 Ser Ala Ser Val Gln Ala Thr Met Tyr Lys Met Ala Glu Gln Ile Leu
 225 230 235 240
 Ala Arg Gln Gln Leu Ile Glu Thr Val Glu Tyr Ser Leu Pro Asn Lys
 245 250 255
 His Tyr Phe Glu Ile Asp Leu Ser Trp His Lys Gly Leu Gln Asn Thr
 260 265 270
 Gly Lys Asn Ala Glu Val Phe Ala Pro Gln Ser Asp Pro Asn Gly Leu
 275 280 285

Ile Lys Cys Thr Val Gly Arg Ser Ser Leu Lys Ser Lys Leu
 290 295 300

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 906 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Preferred sequence for expression in prokaryotes

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGTCTGCGG TAAAAGCAGC GCGCTACGGC AAGGACAATG TTCGCGTCTA CAAGGTTTAC 60
 AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120
 GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTCGC AACCGACTCC 180
 ATTAAGAACA CCATTTACAT CACCGCCAAG CAGAACCCCG TTA CTCTCC CGAGCTGTTC 240
 GGCTCCATCC TGGGCACACA CTTTATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300
 AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360
 TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420
 ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480
 TGGGGCTTCC TGCCTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540
 ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTTCTAGT GACTCCAGGA GGTCCGCTCG 600
 CACGTGCCTA AGTTTCGATG TACCTGGGCC ACTGCTCGCG AGGTCACTCT GAAGACTTTT 660
 GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720
 GCGCGCCAGC AGCTGATCGA GACTGTGAG TACTCGTTGC CTAACAAGCA CTATTTTCGAA 780
 ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840
 CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900
 AAATTG 906

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 906 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Preferred sequence for expression in eukaryotes

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGTCTGCTG TTAAGGCTGC TAGATACGGT AAGGACAACG TTAGAGTCTA CAAGGTTTAC 60
AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120
GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTGCG AACCGACTCC 180
ATTAAGAACA CCATTACAT CACCGCCAAG CAGAACCCCG TTAATCCTCC CGAGCTGTTC 240
GGCTCCATCC TGGGCACACA CTTTATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300
AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360
TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420
ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480
TGGGGCTTCC TCGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540
ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTTCTAGT GACTCCAGGA GGTCCGCTCG 600
CACGTGCCTA AGTTTCGATG TACCTGGGCC ACTGCTCGCG AGGTCCTCT GAAGACTTTT 660
GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720
GCGCGCCAGC AGCTGATCGA GACTGTGCG TACTCGTTGC CTAACAAGCA CTATTTTGAA 780
ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840
CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900
AAATTG 906

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Preferred non-translated 5' sequence for animal cells

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGCTTGCCGC CACT

14

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 906 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Preferred sequence for expression in animal cells

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGTCCGCAG TAAAAGCAGC CCGCTACGGC AAGGACAATG TCCGCGTCTA CAAGGTTTAC 60
AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120
GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTCGC AACCGACTCC 180
ATTAAGAACA CCATTACAT CACCGCCAAG CAGAACCCCG TTAATCCTCC CGAGCTGTTC 240
GGCTCCATCC TGGGCACACA CTTATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300
AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360
TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420
ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480
TGGGGCTTCC TCGGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540
ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTTCAGTG GACTCCAGGA GGTCCGCTCG 600
CACGTGCCTA AGTTCGATGC TACCTGGGCC ACTGCTCGCG AGGTCACTCT GAAGACTTTT 660
GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720
GCGCGCCAGC AGCTGATCGA GACTGTGAG TACTCGTTGC CTAACAAGCA CTATTTTGAA 780
ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840
CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900
AAATTG 906

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: reverse transcription primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCCGGGCC CTTTTTTTTT TTT

23

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product T 17

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Asn Val Gln Val Asp Val Val Glu Gly Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product T 20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Asn Phe Ser Gly Leu Gln Glu Val
1 5

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product T 23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Phe Asp Ala Thr Trp Ala
1 5

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Hydrolysis product T 27

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

His Tyr Phe Glu Ile Asp Leu Ser
1 5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Hydrolysis product T 28

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Hydrolysis product T 29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

His Tyr Phe Glu Ile Asp Leu Ser Trp His Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Hydrolysis product T 31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ser Thr Asn Ser Gln Phe Trp Gly Phe Leu Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Hydrolysis product T 32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly Ser Ile Leu Gly Thr
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Hydrolysis product T 33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly Ser Ile Leu Gly Thr
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product V 1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Tyr Ser Leu Pro Asn Lys His Tyr Phe Glu Ile Asp Leu Ser Trp His
1 5 10 15

Lys

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product V 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Val Thr Leu Lys Thr Phe Ala Glu Asp Asn Ser Ala Ser Val Gln Ala
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product V 3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Thr Ser Tyr Thr Lys Ala Asp Asn Ser Val Ile Val Asp Thr Asp Ser
1 5 10 15

Ile Lys Asn Thr Ile Tyr Ile Thr
20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product V 5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Gly Lys Gly Ile Asp Ile Lys Ser Ser Leu Ser Gly Leu Thr Val Leu
1 5 10 15

Lys Ser Thr Asn Ser Gln Phe Trp Gly Phe Leu Arg
20 25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Hydrolysis product V 6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Gly Lys Gly Ile Asp Ile Lys Ser Ser Leu Ser Gly Leu Thr Val Leu
1 5 10 15

Lys

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1236 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Fragment 3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GATCCGCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA TGAGTGAGCT AACTTACATT 60
AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC AGCTGCATTA 120
ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCCAGG GTGGTTTTTC 180
TTTTCAACCAG TGAGACGGGC AACAGCTGAT TGCCCTTCAC CGCCTGGCCC TGAGAGAGTT 240
GCAGCAAGCG GTCCACGCTG GTTTGCCCCA CCACCCGAAA ATCCTGTTTG ATGGTGGTTA 300
ACGGCGGGAT ATAACATGAG CTGTCTTCGG TATCGTCGTA TCCCACTACC GAGATATCCG 360
CACCAACGCG CAGCCCGGAC TCGGTAATGG CGCGCATTGC GCCCAGCGCC ATCTGATCGT 420
TGGCAACCAG CATCGCAGTG GGAACGATGC CCTCATTAG CATTGTCATG GTTTGTTGAA 480
AACCGGACAT GGCACCTCAG TCGCCTTCCC GTTCCGCTAT CGGCTGAATT TGATTGCGAG 540
TGAGATATTT ATGCCAGCCA GCCAGACGCA GACGCGCCGA GACAGAACTT AATGGGCCCC 600
CTAACAGCGC GATTTGCTGG TGACCCAATG CGACCAGATG CTCCACGCCC AGTCGCGTAC 660
CGTCTTCATG GGAGAAAATA ATACTGTTGA TGGGTGTCTG GTCAGAGACA TCAAGAAATA 720
ACGCCGGAAC ATTAGTGAGC GCAGCTTCCA CAGCAATGGC ATCCTGGTCA TCCAGCGGAT 780
AGTTAATGAT CAGCCCACTG ACGCGTTGCG CGAGAAGATT GTGCACCGCC GCTTTACAGG 840
CTTCGACGCC GCTTCGTTCT ACCATCGACA CCACCACGCT GGCACCCAGT TGATCGGCGC 900
GAGATTTAAT CGCCGCGACA ATTTGCGACG GCGCGTGCAG GGCCAGACTG GAGGTGGCAA 960
CGCCAATCAG CAACGACTGT TTGCCCCGCA GTTGTGTGTC CACGCGGTTG GGAATGTAAT 1020
TCAGCTCCGC CATCGCCGCT TCCACTTTTT CCCGCGTTTT CGCAGAAACG TGGCTGGCCT 1080
GGTTCACCAC GCGGGAAACG GTCTGATAAC AGACACCGGC ATACTCTGCG ACATCGTATA 1140
ACGTTACTGG TTTCACATTC ACCACCCTGA ATTGACTCTC TTCCGGGCGC TATCATGCCA 1200
TACCGCGAAA GGTTTTGCGC CATTGATGG TGTCCG 1236

2
C

(A) LENGTH: 321 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(B) CLONE: Fragment 4

(A) NAME/KEY: CDS

(D) OTHER INFORMATION: /product= "regulatory signal + aa
1-44 human growth hormone precursor"

TCGAGCTGAC	TGACCTGTTG	CTTATATTAC	ATCGATAGCG	TATAATGTGT	GGAATTGTGA	60
GCGATAACAA	TTTCACACAG	TTTAACTTTA	AGAAGGAGAT	ATACAT	ATG GCT ACC	115
					Met Ala Thr	
					1	
GGA TCC CGG ACT AGT CTG CTC CTG GCT TTT GGC CTG CTC TGC CTG CCC						163
Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu Cys Leu Pro						
	5				10	15
TGG CTT CAA GAG GGC AGT GCC TTC CCA ACC ATT CCC TTA TCT AGA CTT						211
Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu						
	20			25	30	35
TTT GAC AAC GCT ATG CTC CGC GCC CAT CGT CTG CAC CAG CTG GCC TTT						259
Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe						
			40		45	50
GAC ACC TAC CAG GAG TTT GAA GAA GCC TAT ATC CCA AAG GAA CAG AAG						307
Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys						
		55			60	65
TAT TCA TTC CTGCA						321
Tyr Ser Phe						
		70				

(A) LENGTH: 70 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met Ala Thr Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu
1 5 10 15

Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu
20 25 30
Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln
35 40 45
Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys
50 55 60
Glu Gln Lys Tyr Ser Phe
65 70

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 74 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: ClaI-NdeI fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGATAGCGTA TAATGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGT TTTTCGCGAA	60
GAAGGAGATA TACA	74

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 190 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Plasmid p373,2 fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GATCTTCAAG CAGACCTACA GCAAGTTCGA CACAACTCA CACAACGATG ACGCACTACT	60
CAAGAACTAC GGGCTGCTCT ACTGCTTCAG GAAGGACATG GACAAGGTCG AGACATTCTT	120
GCGCATCGTG CAGTGCCGCT CTGTGGAGGG CAGCTGTGGC TTCTAGTAAG GTACCCTGCC	180
CTACGTACCA	190

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: AccI-NdeI synthetic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TATGTCTGCG GTAAAAGCAG CGCGCTACGG CAAGGACAAT GTTCGCGT

48

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 360 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Plasmid pEMR469 fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGGACGCGTC TCCTCTGCCG GAACACCGGG CATCTCCAAC TTATAAGTTG GAGAAATAAG	60
AGAATTTTCAG ATTGAGAGAA TGAAAAAAAA AAAAAAAAAA AAGGCAGAGG AGAGCATAGA	120
AATGGGGTTC ACTTTTGGT AAAGCTATAG CATGCCTATC ACATATAAAT AGAGTGCCAG	180
TAGCGACTTT TTTCACACTC GAGATACTCT TACTACTGCT CTCTTGTTGT TTTTATCACT	240
TCTTGTTTCT TCTTGGTAAA TAGAATATCA AGCTACAAAA AGCATACAAT CAACTATCAA	300
CTATTAATA TATCGATACC ATATGGATCC GTCGACTCTA GAGGATCGTC GACTCTAGAG	360

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Fragment C

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CGATATACAC AATGTCTGCT GTTAAGGCTG CTAGATACGG TAAGGACAAC GTTAGAGT 58

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1013 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: Fragment D

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTACAAGGTT CACAAGGACG AGAAGACCGG TGTCCAGACG GTGTACGAGA TGACCGTCTG 60
TGTGCTTCTG GAGGGTGAGA TTGAGACCTC TTACACCAAG GCCGACAACA GCGTCATTGT 120
CGCAACCGAC TCCATTAAGA ACACCATTTA CATCACCGCC AAGCAGAACC CCGTTACTCC 180
TCCCGAGCTG TTCGGCTCCA TCCTGGGCAC ACACTTCATT GAGAAGTACA ACCACATCCA 240
TGCCGCTCAC GTCAACATTG TCTGCCACCG CTGGACCCGG ATGGACATTG ACGGCAAGCC 300
ACACCCTCAC TCCTTCATCC GCGACAGCGA GGAGAAGCGG AATGTGCAGG TGGACGTGGT 360
CGAGGGCAAG GGCATCGATA TCAAGTCGTC TCTGTCCGGC CTGACCGTGC TGAAGAGCAC 420
CAACTCGCAG TTCTGGGGCT TCCTGCGTGA CGAGTACACC ACACTTAAGG AGACCTGGGA 480
CCGTATCCTG AGCACCGACG TCGATGCCAC TTGGCAGTGG AAGAATTTC A GTGGACTCCA 540
GGAGGTCCGC TCGCACGTGC CTAAGTTCGA TGCTACCTGG GCCACTGCTC GCGAGGTCAC 600
TCTGAAGACT TTTGCTGAAG ATAACAGTGC CAGCGTGCAG GCCACTATGT ACAAGATGGC 660
AGAGCAAATC CTGGCGCGCC AGCAGCTGAT CGAGACTGTC GAGTACTCGT TGCCTAACAA 720
GCACTATTTT GAAATCGACC TGAGCTGGCA CAAGGGCCTC CAAAACACCG GCAAGAACGC 780
CGAGGTCTTC GCTCCTCAGT CGGACCCCAA CGGTCTGATC AAGTGTACCG TCGGCCGGTC 840
CTCTCTGAAG TCTAAATTGT AAACCAACAT GATTCTCACG TTCCGGAGTT TCCAAGGCAA 900
ACTGTATATA GTCTGGGATA GGGTATAGCA TTCATTCACT TGTTTTTTTAC TTCAAAAAA 960
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGC CCG 1013

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 207 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

- (vii) IMMEDIATE SOURCE:
 (B) CLONE: Synthetic GAL7 fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGCGTCTATA CTTCGGAGCA CTGTTGAGCG AAGGCTCATT AGATATATTT TCTGTCATTT 60
TCCTTAACCC AAAAATAAGG GAGAGGGTCC AAAAAGCGCT CGGACAACTG TTGACCGTGA 120
TCCGAAGGAC TGGCTATACA GTGTTCAAA AATAGCCAAG CTGAAAATAA TGTGTAGCCT 180
TTAGCTATGT TCAGTTAGTT TGGCATG 207

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

- (vii) IMMEDIATE SOURCE:
 (B) CLONE: Modified XbaI-MluI adapter

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CTAGGCTAGC GGGCCCGCAT GCA 23

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 422 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

- (vii) IMMEDIATE SOURCE:
 (B) CLONE: Plasmid pSE1 "site binding to HindIII"
 fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AGCTGGCTCG CATCTCTCCT TCACGCGCCC GCCGCCCTAC CTGAGGCCCG CATCCACGCC 60
GGTGAGTCGC GTTCTGCCGC CTCCCGCCTG TGGTGCCTCC TGAAGTGCCT CCGCCGTCTA 120
GGTAGGCTCC AAGGGAGCCG GACAAAGGCC CGGTCTCGAC CTGAGCTCTA AACTTACCTA 180
GACTCAGCCG GCTCTCCACG CTTTGCCTGA CCCTGCTTGC TCAACTCTAC GTCTTTGTTT 240
CGTTTTCTGT TCTGCGCCGT TACAATTCA AGGTATGCGC TGGGACCTGG CAGGCGGCAT 300
CTGGGACCCC TAGGAAGGGC TTGGGGGTCC TCGTGCCCAA GGCAGGGAAC ATAGTGGTCC 360
CAGGAAGGGG AGCAGAGGCA TCAGGGTGTC CACTTTGTCT CCGCAGCTCC TGAGCCTGCA 420
GA 422

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 77 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Synthetic HindIII-"site binding to BamHI"
fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGCTTGTCGA CTAATACGAC TCACTATAGG GCGGCCGCGG GCCCCTGCAG GAATTCGGAT 60
CCCCCGGGTG ACTGACT 77

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Synthetic HindIII-AccI fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AGCTTGCCGC CACTATGTCC GCAGTAAAAG CAGCCCGCTA CGGCAAGGAC AATGTCCGCG 60
T 61

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 920 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: HindIII-SnaBI fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

C7
Comit

AGCTTGCCGC	CACTATGTCC	GCAGTAAAG	CAGCCCGCTA	CGGCAAGGAC	AATGTCCGCG	60
TCTACAAGGT	TCACAAGGAC	GAGAAGACCG	GTGTCCAGAC	GGTGTACGAG	ATGACCGTCT	120
GTGTGCTTCT	GGAGGGTGAG	ATTGAGACCT	CTTACACCAA	GGCCGACAAC	AGCGTCATTG	180
TCGCAACCGA	CTCCATTAAG	AACACCATT	ACATCACCGC	CAAGCAGAAC	CCCGTTACTC	240
CTCCCGAGCT	GTTCCGGCTCC	ATCCTGGGCA	CACACTTCAT	TGAGAAGTAC	AACCACATCC	300
ATGCCGCTCA	CGTCAACATT	GTCTGCCACC	GCTGGACCCG	GATGGACATT	GACGGCAAGC	360
CACACCCTCA	CTCCTTCATC	CGCGACAGCG	AGGAGAAGCG	GAATGTGCAG	GTGGACGTGG	420
TCGAGGGCAA	GGGCATCGAT	ATCAAGTCGT	CTCTGTCCGG	CCTGACCGTG	CTGAAGAGCA	480
CCAACCTCGA	GTTCTGGGGC	TTCCTGCGTG	ACGAGTACAC	CACACTTAAG	GAGACCTGGG	540
ACCGTATCCT	GAGCACCGAC	GTCGATGCCA	CTTGGCAGTG	GAAGAATTTC	AGTGGACTCC	600
AGGAGGTCCG	CTCGCACGTG	CCTAAGTTCG	ATGCTACCTG	GGCCACTGCT	CGCGAGGTCA	660
CTCTGAAGAC	TTTTGCTGAA	GATAACAGTG	CCAGCGTGCA	GGCCACTATG	TACAAGATGG	720
CAGAGCAAAT	CCTGGCGCGC	CAGCAGCTGA	TCGAGACTGT	CGAGTACTCG	TTGCCTAACA	780
AGCACTATTT	CGAAATCGAC	CTGAGCTGGC	ACAAGGGCCT	CCAAAACACC	GGCAAGAACG	840
CCGAGGTCTT	CGCTCCTCAG	TCGGACCCCA	ACGGTCTGAT	CAAGTGATACC	GTCGGCCGGT	900
CCTCTCTGAA	GTCTAAATTG					920